Perbandingan *Linear Regression* Dan *Random Forest* Dalam Prediksi Tingkat Kemiskinan Di Indonesia Berdasarkan Provinsi

Comparison of Linear Regression and Random Forest Methods for Predicting Poverty Levels in Indonesia Based on Provincial

Donna Nur Tamara*[1], Shafa Syahida*[2], Arif Mulya Putra*[3], Diah Septiani*[4]
[1,2,3,4] Program Studi S1 Sains Data, Telkom University, Purwokerto, Jawa Tengah
e-mail: *[1]donnanur@student.telkomuniversity.ac.id,
*[2]shafasyahida@student.telkomuniversity.ac.id,
*[3]arifmulyaputra@student.telkomuniversity.ac.id,
*[4]diahseptiani@telkomuniversity.ac.id

Abstrak - *Kemiskinan merupakan salah satu permasalahan utama dalam pembangunan nasional yang dipengaruhi oleh berbagai faktor sosial dan ekonomi. Penelitian ini bertujuan membandingkan kinerja metode Linear Regression dan Random Forest Regression dalam memprediksi tingkat kemiskinan provinsi di Indonesia periode 2026-2025. Tahapan penelitian meliputi EDA, preprocessing, rekayasa fitur, serta pembagian data. Model dievaluasi menggunakan metrik MSE, RMSE, MAE, dan R^2. Hasil penelitian menunjukkan bahwa kedua model berkinerja sangat baik dengan nilai R^2 di atas 0,99. Namun, Linear Regression memberikan performa terbaik secara konsisten dengan nilai sebesar 99,68% (MSE 0,1764) dibandingkan Random Forest sebesar 99,65% (MSE 0,1968), mengindikasikan pola hubungan antar-variabel makro terhadap kemiskinan pada dataset ini cenderung linier. Hasil prediksi periode 2026-2028 menunjukkan tren penurunan rata-rata kemiskinan nasional dari 10,20% menjadi 9,69%. Hasil penelitian ini diharapkan menjadi rujukan ilmiah bagi pemerintah dalam merumuskan kebijakan penanganan kemiskinan yang tepat sasaran.*

Kata kunci – kemiskinan; machine learning; prediksi; random forest; regresi linear.

Abstract *- Poverty remains a fundamental challenge in national development, influenced by various socioeconomic factors. This study aims to compare the performance of Linear Regression and Random Forest Regression methods in predicting provincial poverty levels in Indonesia for the 2010-2025 period. The research stages encompass EDA, preprocessing, feature engineering, and data splitting. The models were evaluated using MSE, RMSE, MAE, and R^2 metrics. The results indicate that both models perform exceptionally well, achieving R^2 scores above 0.99. However, Linear Regression consistently delivers superior performance, yielding an R^2 of 99.68% (MSE 0.1764) compared to Random Forest's 99.65% (MSE 0.1968), which implies that the relationships between the macroeconomic variables and poverty in this dataset tend to be linear. Forecasting results for the 2026-2028 period project a declining trend in the national average poverty rate, moving from 10.20% to 9.69%. These projections are expected to serve as a scientific reference for the government in formulating well-targeted poverty alleviation strategies.*

Keywords – poverty; machine learning; prediction; random forest; Linear Regression.

I. PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan utama yang masih dihadapi oleh banyak negara berkembang, termasuk Indonesia. Tingkat kemiskinan menjadi indikator penting dalam mengukur keberhasilan pembangunan suatu negara karena berkaitan langsung dengan kesejahteraan masyarakat. Menurut data Badan Pusat Statistik (BPS), meskipun terjadi penurunan tingkat kemiskinan dalam beberapa tahun terakhir, kesenjangan antar wilayah, khususnya antar provinsi, masih menjadi tantangan yang signifikan. Hal ini menunjukkan bahwa distribusi kesejahteraan belum merata dan memerlukan pendekatan yang lebih efektif dalam perencanaan kebijakan [1]. Berbagai faktor mempengaruhi tingkat kemiskinan, di antaranya adalah tingkat pendidikan, kesehatan, pengangguran, serta kondisi ekonomi daerah. Kompleksitas hubungan antar variabel tersebut menyebabkan kesulitan dalam melakukan prediksi menggunakan pendekatan konvensional. Oleh karena itu, diperlukan metode yang mampu menangkap hubungan yang kompleks dan non-linear antar variabel untuk menghasilkan prediksi yang lebih akurat [2].

Seiring dengan perkembangan teknologi, pendekatan berbasis *machine learning* telah banyak digunakan dalam berbagai bidang, termasuk ekonomi dan sosial. *Machine learning* mampu memproses data dalam jumlah besar serta menemukan pola tersembunyi yang tidak mudah diidentifikasi dengan metode statistik tradisional. Penelitian terbaru menunjukkan bahwa metode *machine learning* memiliki performa yang lebih baik dibandingkan metode konvensional dalam berbagai kasus prediksi, termasuk prediksi kemiskinan [3]. Beberapa penelitian sebelumnya telah menggunakan berbagai metode *machine learning* untuk memprediksi tingkat kemiskinan. Misalnya, penelitian oleh Zhang dkk. [4] menunjukkan bahwa model berbasis *ensemble* seperti Random Forest mampu memberikan hasil prediksi yang lebih akurat dibandingkan model linear. Selain itu, penelitian oleh Kumar dan Singh [5] juga menyatakan bahwa metode Random Forest memiliki keunggulan dalam menangani data dengan hubungan non-linear dan variabilitas tinggi. Namun demikian, metode Regresi Linier masih banyak digunakan karena kemudahannya dalam interpretasi dan implementasi.

Meskipun banyak penelitian telah dilakukan, terdapat beberapa keterbatasan yang masih perlu dikaji lebih lanjut. Pertama, sebagian besar penelitian hanya berfokus pada satu metode tanpa melakukan perbandingan antara metode linear dan non-linear. Kedua, penelitian sebelumnya seringkali tidak mengoptimalkan performa model melalui proses pemilihan fitur dan *tuning hyperparameter*, padahal kedua proses tersebut sangat berpengaruh terhadap akurasi model [6]. Ketiga, kajian yang secara khusus menganalisis prediksi kemiskinan pada tingkat provinsi di Indonesia masih relatif terbatas. Berdasarkan permasalahan tersebut, penelitian ini berupaya untuk melakukan perbandingan antara metode Regresi Linier dan Random Forest dalam memprediksi tingkat kemiskinan di Indonesia berdasarkan data provinsi. Selain itu, penelitian ini juga mengintegrasikan proses *feature selection* dan *hyperparameter tuning* untuk meningkatkan performa model. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam bentuk pendekatan yang lebih komprehensif dalam pemodelan prediksi kemiskinan.

Gap dalam penelitian ini terletak pada perbandingan dua pendekatan yang berbeda, yaitu metode linear dan metode berbasis *ensemble*, dengan mempertimbangkan optimasi model melalui *feature selection* dan *hyperparameter tuning*. Selain itu, penelitian ini juga memberikan fokus pada analisis tingkat provinsi, yang memungkinkan identifikasi pola kemiskinan secara lebih spesifik dibandingkan penelitian sebelumnya. Dengan demikian, penelitian ini memiliki

nilai kebaruan dalam hal pendekatan metodologi dan ruang lingkup analisis.
Permasalahan utama yang diangkat dalam penelitian ini adalah bagaimana membandingkan performa metode Regresi Linier dan Random Forest dalam memprediksi tingkat kemiskinan di Indonesia, serta bagaimana pengaruh *feature selection* terhadap peningkatan hasil prediksi model. Penelitian ini dilakukan dengan tujuan untuk membandingkan performa metode Regresi Linier dan Random Forest dalam prediksi tingkat kemiskinan, menganalisis pengaruh *feature selection* dan *hyperparameter tuning* terhadap performa model, serta menghasilkan model prediksi yang dapat digunakan untuk memperkirakan tingkat kemiskinan pada beberapa tahun ke depan di tingkat provinsi. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi bagi pengambil kebijakan dalam merumuskan strategi penanganan kemiskinan yang lebih efektif dan berbasis data.

II. TINJAUAN TEORI

1. Kemiskinan

Kemiskinan merupakan kondisi ketidakmampuan individu atau kelompok dalam memenuhi kebutuhan dasar seperti pangan, sandang, dan papan. Tingkat kemiskinan sering digunakan sebagai indikator utama dalam mengukur kesejahteraan masyarakat suatu negara. Faktor-faktor yang mempengaruhi kemiskinan meliputi tingkat pendidikan, kesehatan, kesempatan kerja, serta kondisi ekonomi secara umum. Di Indonesia, pengukuran kemiskinan umumnya dilakukan berdasarkan garis kemiskinan yang ditetapkan oleh Badan Pusat Statistik [7].

2. Regresi Linier

Machine learning merupakan cabang dari kecerdasan buatan yang memungkinkan sistem untuk belajar dari data dan membuat prediksi tanpa diprogram secara eksplisit. Dalam konteks prediksi, *machine learning* termasuk dalam kategori *supervised learning*, di mana model dilatih menggunakan data historis yang memiliki label. Pendekatan ini banyak digunakan dalam berbagai bidang, termasuk ekonomi dan sosial, karena kemampuannya dalam menangkap pola kompleks dalam data [8]. Persamaan dari model Regresi Linear sebagai berikut:

$$y = a + bx \quad (1)$$

Menghitung konstanta (a) dapat dilihat pada persamaan (2) sebagai berikut:

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \quad (2)$$

Menghitung koefisien (b) dapat dilihat pada persamaan (3) sebagai berikut:

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (3)$$

Pada Persamaan (1), (2), dan (3), y merupakan variabel dependen (variabel terikat), x merupakan variabel independen (variabel bebas), a merupakan konstanta atau nilai y ketika $x = 0$, dan b merupakan koefisien regresi yang menunjukkan besarnya perubahan pada variabel dependen akibat perubahan satu satuan pada variabel independent [9].

3. Random Forest

Random Forest merupakan metode *ensemble learning* yang menggabungkan banyak *decision tree* untuk meningkatkan akurasi prediksi dan mengurangi *overfitting*. Setiap *tree* dibangun menggunakan subset data yang berbeda dan hasil prediksi akhir diperoleh melalui proses agregasi. Metode ini mampu menangkap hubungan non-linear dan interaksi antar variabel dengan baik, sehingga sering digunakan dalam berbagai permasalahan prediksi. Rumus umum dari model Random Forest sebagai berikut [10]:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (4)$$

Pada Persamaan (4), $\hat{y}$ merupakan prediksi akhir yang dihasilkan oleh model Random Forest Regression, T menyatakan jumlah pohon (*trees*) yang digunakan dalam proses pembentukan model, dan $h_t(x)$ merupakan hasil prediksi yang diberikan oleh pohon ke-t terhadap data masukan x. Prediksi akhir diperoleh dengan menghitung rata-rata seluruh hasil prediksi yang dihasilkan oleh setiap pohon dalam hutan (*forest*), sehingga metode ini mampu menghasilkan prediksi yang lebih stabil dan mengurangi risiko *overfitting* dibandingkan penggunaan satu *decision tree* tunggal [10].

4. *Feature Engineering* dan *Lag Feature*

Feature engineering merupakan proses transformasi, ekstraksi, dan pembentukan fitur baru dari data asli untuk meningkatkan kualitas representasi data yang digunakan oleh model *Machine Learning*. Proses ini bertujuan menghasilkan fitur yang lebih informatif sehingga pola yang terkandung dalam data dapat dipelajari secara lebih efektif dan menghasilkan performa prediksi yang lebih baik. Berbagai penelitian menunjukkan bahwa kualitas fitur memiliki pengaruh yang signifikan terhadap akurasi model, bahkan sering kali lebih menentukan dibandingkan pemilihan algoritma yang digunakan [11].

Pada penelitian ini dilakukan beberapa tahapan *feature engineering*, yaitu transformasi logaritmik pada variabel Produk Domestik Regional Bruto (PDRB) menjadi Log_PDRB, pengkodean variabel semester dan provinsi, serta pembentukan fitur historis (*lag feature*). *Lag feature* merupakan fitur yang dibentuk dari nilai variabel pada periode sebelumnya dan banyak digunakan dalam permasalahan peramalan (*forecasting*) untuk menangkap pola temporal pada data runtun waktu. Penggunaan fitur Kemiskinan_lag1 dan Kemiskinan_lag2 bertujuan untuk merepresentasikan pengaruh tingkat kemiskinan pada periode sebelumnya terhadap tingkat kemiskinan pada periode saat ini, sehingga model dapat mempelajari ketergantungan antar waktu secara lebih baik [11].

5. Evaluasi Model Regresi

Evaluasi model dilakukan untuk mengukur tingkat akurasi prediksi yang dihasilkan. Beberapa metrik yang digunakan dalam penelitian ini adalah:

***Mean Squared Error* (MSE)**

Mean Squared Error (MSE) digunakan untuk mengukur rata-rata kuadrat selisih antara nilai aktual dan nilai prediksi. Nilai MSE yang lebih kecil menunjukkan performa model yang lebih baik [12].

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2 \quad (5)$$

Pada Persamaan (5), m menyatakan jumlah observasi, X_i merupakan nilai aktual pada observasi ke- i, dan Y_i merupakan nilai prediksi pada observasi ke- i.

***Root Mean Squared Error* (RMSE)**

RMSE merupakan akar kuadrat dari MSE yang digunakan untuk mengembalikan satuan kesalahan ke satuan asli data sehingga lebih mudah diinterpretasikan [12].

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \quad (6)$$

Pada Persamaan (6), m menyatakan jumlah observasi data, X_i merupakan nilai aktual tingkat kemiskinan pada observasi ke-i, dan Y_i merupakan nilai prediksi tingkat kemiskinan pada observasi ke- i.

***Mean Absolute Error* (MAE)**

MAE mengukur rata-rata nilai absolut dari selisih antara nilai aktual dan nilai prediksi. Semakin kecil nilai MAE menunjukkan kesalahan prediksi yang semakin rendah [12].

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|X_i - Y_i| \quad (7)$$

Pada Persamaan (7), m menyatakan jumlah observasi data, X_i merupakan nilai aktual tingkat kemiskinan pada observasi ke-i, dan Y_i merupakan nilai prediksi tingkat kemiskinan pada observasi ke-i.

***Coefficient of Determination* (R^2)**

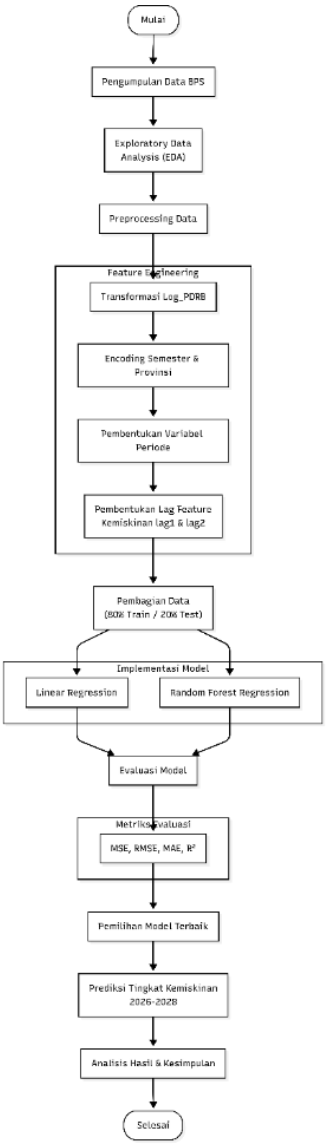
Koefisien determinasi (R^2) digunakan untuk mengukur kemampuan model dalam menjelaskan variasi data aktual. Nilai R^2 berada pada rentang $(-\infty,1]$, di mana nilai yang semakin mendekati 1 menunjukkan model memiliki kemampuan prediksi yang semakin baik [12].

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(X_i - Y_i)^2}{\sum_{i=1}^{m}(\bar{Y} - Y_i)^2} \quad (8)$$

Pada Persamaan (8), X_i merupakan nilai aktual tingkat kemiskinan pada observasi ke- , i y_i merupakan nilai prediksi tingkat kemiskinan pada observasi ke- , i $\bar{Y}$ menyatakan nilai rata-rata (mean) dari keseluruhan nilai aktual tingkat kemiskinan, dan menyatakan jumlah observasi data. m

III. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis data berbasis *data mining* untuk memodelkan dan memprediksi hubungan antar variabel. Metodologi penelitian disusun secara sistematis mulai dari pengumpulan data, tahap *preprocessing*, pemodelan, hingga evaluasi model.



Gambar 1. Alur Penelitian

Berdasarkan Gambar 1 di atas, penjabaran lebih detail mengenai alur penelitian dalam penelitian ini sebagai berikut:

1. Pengumpulan Data

Tahap pertama adalah pengumpulan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Indonesia. Data yang digunakan merupakan data tingkat kemiskinan provinsi di Indonesia periode 2010-2025 yang disusun dalam bentuk data dengan periode semesteran. Selain tingkat kemiskinan sebagai variabel target, penelitian ini juga menggunakan beberapa variabel prediktor, yaitu Indeks Pembangunan Manusia (IPM), Rasio Gini, Tingkat Pengangguran Terbuka (TPT), dan Produk Domestik Regional Bruto (PDRB).

2. *Exploratory Data Analysis* (EDA) dan *Preprocessing Data*

Setelah data terkumpul, dilakukan tahap *Exploratory Data Analysis* (EDA) untuk memahami karakteristik data sebelum proses pemodelan. Analisis yang dilakukan meliputi pemeriksaan distribusi data, identifikasi hubungan antar variabel menggunakan matriks

evaluasi, serta visualisasi tren tingkat kemiskinan berdasarkan tahun dan provinsi di Indonesia.

Selanjutnya dilakukan tahap *preprocessing* untuk meningkatkan kualitas data. Proses ini meliputi penanganan data yang hilang (*missing values*), penghapusan data yang tidak relevan, pengurutan data berdasarkan provinsi dan periode waktu, serta penyesuaian format data agar sesuai dengan kebutuhan analisis.

3. Feature Engineering

Feature engineering dilakukan untuk meningkatkan kualitas data sehingga model dapat mempelajari pola dengan lebih baik. Pada tahap ini dilakukan transformasi logaritmik terhadap variabel PDRB menjadi Log_PDRB untuk mengurangi tingkat kemencengan data (*skewness*). Selain itu dilakukan pengkodean (*encoding*) pada variabel semester dan provinsi agar dapat diproses oleh algoritma *Machine Learning*.

Penelitian ini juga membentuk fitur waktu berupa variabel Periode dan fitur historis (*lag feature*) yang terdiri dari Kemiskinan_lag1 dan Kemiskinan_lag2. Fitur historis tersebut digunakan untuk menangkap pengaruh tingkat kemiskinan pada periode sebelumnya terhadap tingkat kemiskinan pada periode saat ini.

4. Pembagian Data (*Training* dan *Testing*)

Dataset yang telah melalui proses *preprocessing* dan *feature engineering* kemudian dibagi menjadi data latih (*training set*) dan data uji (*testing set*). Proporsi data yang digunakan adalah 80% sebagai data latih dan 20% sebagai data uji. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi kemampuan model dalam melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya.

5. Penerapan Model (*Linear Regression* dan *Random Forest Regression*)

Pada tahap ini digunakan dua model regresi yaitu Regresi Linier yang digunakan untuk memodelkan hubungan linier antara variabel independen dan tingkat kemiskinan, metode ini dipilih karena memiliki interpretasi yang sederhana dan mampu menunjukkan pengaruh terhadap masing-masing variabel terhadap target. Kemudian, *Random Forest Regression* digunakan sebagai model pembanding karena mampu menangkap hubungan non-linier dan kompleks antar variabel melalui sejumlah *decision tree* yang dibangun menggunakan Teknik *ensemble learning* dalam data. Kedua model dilatih menggunakan data latih yang sama untuk memastikan perbandingan yang adil.

6. Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja dari kedua model yang digunakan, yaitu regresi linier dan *Random Forest regression*. Dalam penelitian ini digunakan empat metrik evaluasi, yaitu *Mean Squared Error* (MSE) digunakan untuk mengukur rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual. Nilai MSE yang lebih kecil menunjukkan performa model yang lebih baik. *Root Mean Squared Error* (RMSE) merupakan akar dari MSE yang memberikan interpretasi dalam satuan yang sama dengan data asli, sehingga lebih mudah dipahami. *Mean Absolute Error* (MAE) digunakan untuk mengukur rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual, tanpa memperhatikan arah kesalahan. Koefisien Determinasi (R^2) digunakan untuk mengukur seberapa besar variansi variabel dependen dapat dijelaskan oleh model. Nilai R^2 yang mendekati 1 menunjukkan model yang semakin baik. Keempat metrik ini digunakan untuk memberikan evaluasi terhadap performa model dalam melakukan prediksi. Model terbaik akan dipilih

berdasarkan hasil evaluasi yang diperoleh dari data uji. Pemilihan dilakukan dengan mempertimbangkan nilai MSE, RMSE, MAE, dan R^2 tertinggi mendekati nilai 1 akan dipilih sebagai model akhir penelitian .

7. Prediksi Tingkat Kemiskinan Tahun 2026-2028

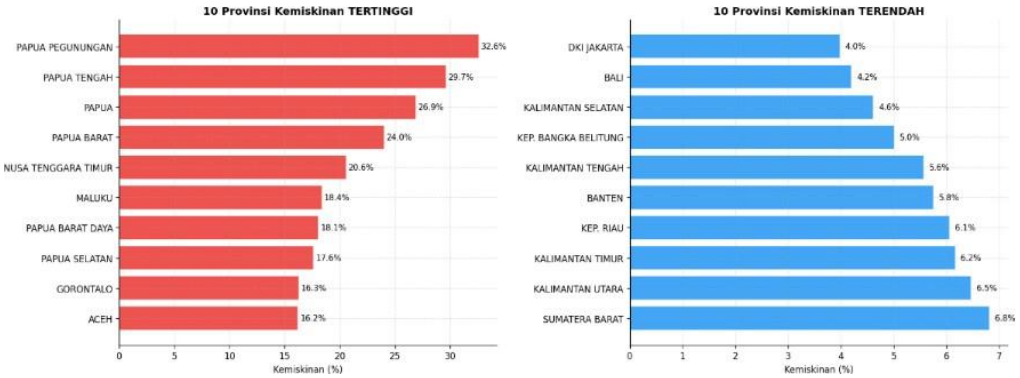
Model terbaik yang diperoleh kemudian digunakan untuk memprediksi tingkat kemiskinan seluruh provinsi di Indonesia pada periode 2026-2028. Proses prediksi dilakukan secara bertahap menggunakan pendekatan *recursive forecasting*, yaitu hasil prediksi pada periode sebelumnya digunakan sebagai masukan (*input*) untuk memprediksi periode berikutnya.

8. Analisis Hasil

Tahap terakhir adalah analisis hasil prediksi yang diperoleh dari model terbaik. Analisis dilakukan untuk mengidentifikasi tren tingkat kemiskinan nasional maupun antar provinsi hingga tahun 2028. Hasil analisis kemudian digunakan sebagai dasar dalam penyusunan kesimpulan penelitian dan rekomendasi bagi pengambilan kebijakan terkait pengentasan kemiskinan.

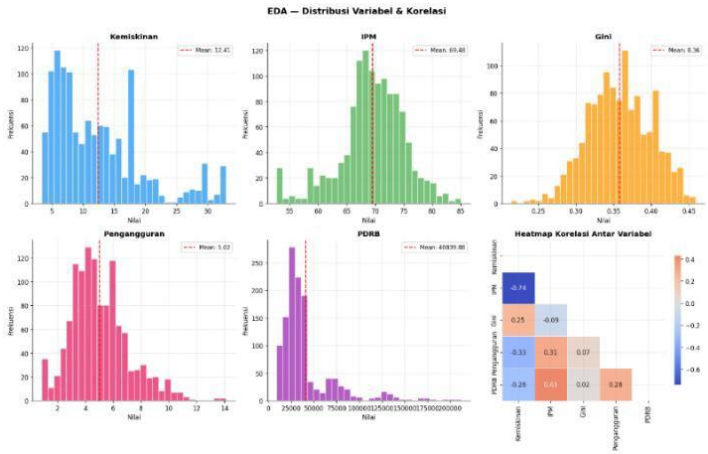
IV. HASIL DAN PEMBAHASAN

1. *Hasil Exploratory Data Analysis* (EDA)



Gambar 2. Top 10 Provinsi dengan Kemiskinan Tertinggi dan Terendah

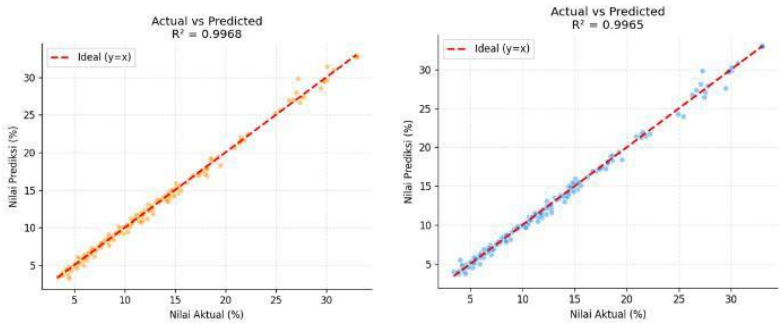
Berdasarkan Gambar 2, terlihat disparitas spasial yang signifikan di mana wilayah Indonesia Timur mendominasi sepuluh provinsi dengan tingkat kemiskinan tertinggi, dipimpin oleh Papua Pegunungan (32,6%), Papua Tengah (29,7%), dan Papua (26,9%). Sebaliknya, sepuluh provinsi dengan tingkat kemiskinan terendah didominasi oleh pusat perekonomian dan industri komoditas, dengan DKI Jakarta mencatatkan angka terkecil secara nasional (4,0%), diikuti oleh Bali (4,2%) dan Kalimantan Selatan (4,6%). Kesenjangan yang kontras antar-wilayah ini mengonfirmasi adanya pengaruh kewilayahan yang kuat terhadap pola kemiskinan di Indonesia, sekaligus menegaskan pentingnya keterlibatan fitur provinsi dalam mengoptimalkan akurasi model prediksi.



Gambar 3. Distribusi Variabel & Korelasi

Berdasarkan Gambar 3, distribusi variabel kemiskinan menunjukkan variasi yang cukup besar antar provinsi dengan rata-rata sebesar 12,41%, sedangkan IPM memiliki rata-rata 69,48 dan cenderung terdistribusi mendekati normal. Variabel Gini memiliki rata-rata sebesar 0,36 dan pengangguran sebesar 5,02%, sementara PDRB menunjukkan distribusi yang menceng ke kanan (*right-skewed*) yang mengindikasikan adanya perbedaan tingkat aktivitas ekonomi yang cukup tinggi antar provinsi. Hasil analisis korelasi menunjukkan bahwa IPM memiliki hubungan negatif yang kuat terhadap tingkat kemiskinan dengan koefisien korelasi sebesar 0,74, yang berarti peningkatan kualitas pembangunan manusia cenderung diikuti oleh penurunan tingkat kemiskinan. Sebaliknya, pengangguran memiliki korelasi positif sebesar 0,33 terhadap kemiskinan, sedangkan variabel Gini dan PDRB menunjukkan hubungan yang relatif lemah terhadap tingkat kemiskinan. Temuan ini menunjukkan bahwa kualitas pembangunan manusia merupakan faktor yang paling berkaitan dengan perubahan tingkat kemiskinan pada data yang digunakan dalam penelitian ini.

2. Hasil Pemodelan



Gambar 4. Perbandingan nilai *Actual* dan *Predicted* pada kedua model

Berdasarkan Gambar 4, kedua model menunjukkan kemampuan prediksi yang sangat baik, yang ditunjukkan oleh sebaran titik yang berada di sekitar garis ideal $(y = x)$. Model *Linear Regression* menghasilkan nilai koefisien determinasi (R^2)sebesar 0,9968, sedikit lebih tinggi

dibandingkan *Random Forest Regression* yang memperoleh nilai (R^2)sebesar 0,9965. Hasil ini menunjukkan bahwa kedua model mampu menjelaskan lebih dari 99% variasi data tingkat kemiskinan, namun *Linear Regression* memiliki tingkat akurasi yang sedikit lebih baik karena menghasilkan prediksi yang lebih mendekati nilai aktual. Temuan ini juga mengindikasikan bahwa hubungan antara variabel prediktor dan tingkat kemiskinan pada dataset penelitian cenderung bersifat linier sehingga dapat dimodelkan dengan baik oleh *Linear Regression*.

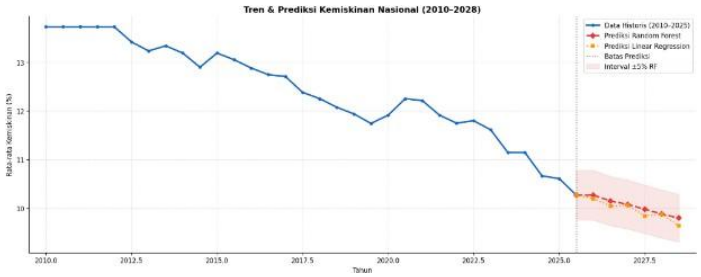
Penelitian ini membandingkan dua algoritma regresi yaitu *Linear Regression* dan Random Forest Regression. Evaluasi dilakukan menggunakan metrik *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi (R^2).

Tabel 1. Perbandingan Kinerja Model

Metrik	Random Forest	*Linear Regression*
MSE	0,1968	0,1764
RMSE	0,4437	0,4200
MAE	0,2868	0,2832
R^2	0,9965	0,9968

Berdasarkan Tabel 1, model *Linear Regression* menghasilkan MSE, RMSE, dan MAE yang lebih rendah dibandingkan Random Forest. Selain itu, nilai R^2 pada *Linear Regression* mencapai 0,9968 yang sedikit lebih unggul dibandingkan dengan Random Forest sebesar 0,9965. Hal ini menunjukkan bahwa model *Linear Regression* mampu menjelaskan lebih dari 99% variasi tingkat kemiskinan pada dataset yang digunakan. Sehingga, model *Linear Regression* dipilih sebagai model terbaik dalam penelitian ini. Meskipun Random Forest dikenal mampu menangkap hubungan non-linear yang kompleks, hasil penelitian menunjukkan bahwa hubungan antara variabel prediktor dan tingkat kemiskinan pada dataset ini cenderung bersifat linier. Kondisi tersebut menyebabkan *Linear Regression* mampu menghasilkan performa yang lebih baik. Tetapi, perbedaan performa kedua model sebenarnya relatif kecil. Hal ini terlihat dari nilai R^2 yang sama-sama berada di atas 0,99. Namun demikian, *Linear Regression* tetap unggul pada seluruh metrik evaluasi sehingga dipilih sebagai model terbaik.

3. Prediksi Tingkat Kemiskinan Tahun 2026-2028



Gambar 5. Tren & Prediksi Kemiskinan Nasional (2010-2028)

Berdasarkan Gambar 5, rata-rata tingkat kemiskinan nasional menunjukkan tren menurun selama periode 2010-2025, dari sekitar 13,7% menjadi 10,3%. Hasil prediksi untuk periode 2026-2028 menunjukkan bahwa tren penurunan tersebut diperkirakan masih berlanjut. Prediksi menggunakan model *Linear Regression* dan Random Forest menghasilkan pola yang relatif serupa, dengan tingkat kemiskinan nasional diproyeksikan berada di bawah 10% pada tahun 2028. Selain itu, interval prediksi yang relatif sempit menunjukkan bahwa model

memiliki tingkat konsistensi yang baik dalam menghasilkan prediksi. Temuan ini mengindikasikan bahwa peningkatan pembangunan manusia, pertumbuhan ekonomi, dan perbaikan kondisi sosial ekonomi berpotensi mendukung penurunan tingkat kemiskinan nasional pada beberapa tahun mendatang.

Model terbaik kemudian digunakan untuk melakukan prediksi tingkat kemiskinan pada seluruh provinsi di Indonesia untuk periode 2026-2028.

Tabel 2. Ringkasan Prediksi Kemiskinan Nasional

Periode	Rata-rata Kemiskinan (%)
2026 Semester 1	10,20
2026 Semester 2	10,02
2027 Semester 1	10,06
2026 Semester 2	9,84
2026 Semester 1	99,88
2026 Semester 2	9,69

Berdasarkan Tabel 2, Hasil prediksi menunjukkan adanya tren penurunan tingkat kemiskinan nasional hingga tahun 2028. Pada Semester 1 tahun 2026 rata-rata tingkat kemiskinan diprediksi sebesar 10,20%, kemudian menurun menjadi 9,69% pada Semester 2 tahun 2028. Tren ini mengindikasikan bahwa kondisi sosial ekonomi Indonesia diperkirakan akan terus membaik apabila pertumbuhan pembangunan manusia, peningkatan ekonomi daerah, dan penurunan pengangguran dapat dipertahankan.

Tabel 3. Lima Provinsi dengan Prediksi Kemiskinan Tertinggi Tahun 2028

No	Provinsi	Prediksi (%)
1	Papua Pegunungan	29,47
2	Papua Tengah	26,28
3	Papua Barat	19,03
4	Papua Selatan	17,93
5	Papua	17,64

Berdasarkan Tabel 3, hasil dari prediksi menunjukkan bahwa wilayah Papua masih diprediksi menghadapi tantangan kemiskinan yang lebih besar dibandingkan provinsi lainnya. Kondisi ini dapat disebabkan oleh faktor geografis, keterbatasan infrastruktur, serta kesenjangan pembangunan antar wilayah.

V. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa model prediksi tingkat kemiskinan di Indonesia berhasil dibangun menggunakan algoritma *Linear Regression* dan *Random Forest Regression* dengan memanfaatkan variabel IPM, Rasio Gini, Tingkat Pengangguran Terbuka (TPT), PDRB, serta fitur historis kemiskinan (*lag feature*). Hasil evaluasi menunjukkan bahwa kedua model memiliki performa yang sangat baik dengan nilai koefisien determinasi (R^2) di atas 0,99. Namun, *Linear Regression* memberikan performa terbaik dengan nilai MSE sebesar 0,1764, RMSE sebesar 0,4200, MAE sebesar 0,2832, dan R^2 sebesar 0,9968, sehingga dipilih sebagai model terbaik dalam penelitian ini. Hasil prediksi menunjukkan bahwa rata-rata tingkat kemiskinan nasional diperkirakan akan terus mengalami penurunan pada periode 2026-2028. Temuan ini mengindikasikan bahwa peningkatan kualitas pembangunan manusia, pertumbuhan ekonomi, dan perbaikan kondisi sosial ekonomi berpotensi mendukung upaya pengurangan kemiskinan di Indonesia pada masa mendatang.

Berdasarkan hasil penelitian tersebut, penelitian selanjutnya disarankan untuk menambahkan variabel sosial ekonomi lainnya agar model dapat menggambarkan faktor-faktor yang memengaruhi kemiskinan dan dapat membandingkan performa model dengan algoritma lainnya. Hasil penelitian ini juga diharapkan dapat menjadi informasi pendukung bagi pemerintah dalam merumuskan kebijakan pengentasan kemiskinan yang lebih tepat sasaran, terutama pada wilayah yang masih memiliki tingkat kemiskinan relatif tinggi.

VI. DAFTAR PUSTAKA

[1] D. B. Wiranatakusuma and G. Primambudi, "Determinants of Poverty in Indonesia," Sociología y Tecnociencia, vol. 11, no. 2, pp. 243-267, 2021.

[2] D. E. Nurjati, "The Socioeconomic Determinants of Poverty Dynamics in Indonesia," MIMBAR: Jurnal Sosial dan Pembangunan, vol. 37, no. 2, pp. 420-431, 2021. S. Kharisma, S.

[3] S. Kharisma, S. P. Maulida, and Y. Caesarani, "Determinant of Poverty Rates in Indonesia: A Panel Data Regression Analysis," Journal of Policy and Sustainable Development Studies, vol. 2, no. 3, pp. 1-12, 2025.

[4] I. Wardhana, "Determinants of Poverty in Central Kalimantan, Indonesia: Evidence from Panel Data Analysis," GROWTH: Journal Magister Ilmu Ekonomi, vol. 11, no. 2, pp. 89-101, 2025.

[5] W. Fhaldian and A. Fahmi, "Explainable Machine Learning for Poverty Prediction in Central Java Regencies and Cities," Sinkron: Jurnal dan Penelitian Teknik Informatika, vol. 9, no. 4, pp. 2500-2512, 2024.

[6] S. Bhahri, Renny, and Rachmat, "Machine Learning for Predicting Poverty and Educational Outcomes: A Comparative Simulation Study for Evidence-Based Social Policy," Journal of Embedded Systems, Security and Intelligent Systems, vol. 6, no. 1, pp. 45-59, 2025.

[7] I. W. Prastiyo and A. Febriandirza, "Analisis Perbandingan Prediksi Tingkat Kemiskinan Menggunakan Metode XGBoost dan Random Forest Regression," Jurnal Media Informatika Budidarma, vol. 8, no. 3, pp. 1694-1702, 2024.

[8] M. Muth, M. Lingenfelder, and G. Nufer, "The application of machine learning for demand prediction under macroeconomic volatility: a systematic literature review," Management Review Quarterly, vol. 75, pp. 2759-2802, 2025.

[9] I. G. Ritonga, Hartono, and P. S. P. Sitorus, "Analisis regresi linear terhadap tingkat pengangguran terbuka menurut provinsi periode 2021-2025," Juktisi: Jurnal Komputer Teknologi Informasi Sistem Komputer, vol. 4, no. 3, pp. 2217-2223, Feb. 2026.

[10] M. B. S. Qolbi, T. N. Puteh, Rivandi, and C. Rozikin, "Prediksi harga rumah di Jakarta Pusat menggunakan algoritma machine learning," Jurnal Ilmu Komputer dan Bisnis (JIKB), vol. 16, no. 1, pp. 16-24, Mei 2025.

[11] I. W. Prastiyo and A. Febriandirza, "Analisis Perbandingan Prediksi Tingkat Kemiskinan Menggunakan Metode XGBoost dan Random Forest Regression," Jurnal Media Informatika Budidarma, vol. 8, no. 3, pp. 1694-1702, 2024.

[12] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Comput. Sci., vol. 7, p. e623, Jul. 2021.